AMERICAN JOURNAL OF PHYSICAL ANTHROPOLOGY 106:547–552 (1998)

Technical Note: Different Techniques, Different Results—A Comparison of Photogrammetric and Caliper-Derived Measurements

DEBORAH J. SHANER,[1] J. STEPHEN BAMFORTH,[2] ARTHUR E. PETERSON,[3] AND OWEN B. BEATTIE[1]*
[1]*Department of Anthropology, University of Alberta, Edmonton, Alberta, Canada T6G 1H4*
[2]*Department of Medical Genetics, University of Alberta, Edmonton, Alberta, Canada T6G 2B7*
[3]*Department of Civil Engineering, University of Alberta, Edmonton, Alberta, Canada T6G 2G7*

KEY WORDS facial features; marked landmarks; unmarked landmarks; measurement variability

ABSTRACT The primary goal of our study was to compare photogrammetric measurements with caliper-derived measurements. We also looked at the difference between caliper-derived measurements that were taken with and without the landmarks marked. Thirteen facial measurements were repeated ten times on two adult subjects as follows: 1) Calipers were used to take the measurements before the landmarks were marked on each subject's face; 2) the landmarks were then marked with a black pencil, and the calipers were used to take the measurements again; and 3) images were taken of each subject with the markings left on the face, and the measurements were extracted from these images. Compared with the caliper-derived data taken with the landmarks marked, the photogrammetric means and standard deviations were typically larger, leading us to conclude that there was a systematic difference between the data. The generally greater variation in the photogrammetric measurements was ascribed to poor conditions, such as shadows, oblique markings, and unmarked landmarks. When the data gathered by caliper with and without the landmarks marked were compared, a systematic difference was suggested by the number of statistically significant t-test probabilities. Marking the landmarks reduced the standard deviations in some measurements by controlling two sources of variation: differing pressure on the skin and slippage of the calipers. Anthropologists, medical geneticists, and others who use measurements for diagnostic or classificatory purposes should be aware that data gathered by different techniques may yield different results. Am J Phys Anthropol 106:547–552, 1998.

Allanson et al. (1993) reported that up to 30 minutes were needed to take 21 facial measurements with calipers on cooperative subjects with Down syndrome. We were planning to take facial measurements from children with genetic syndromes but wanted to use a method that was faster than calipers, in the hopes that participation rates would be higher. Photogrammetry, a method by which measurements are extracted from images, seemed to be a good choice, because

*Correspondence to: Owen B. Beattie, Department of Anthropology, 13–15 HM Tory Building, University of Alberta, Edmonton, Alberta, Canada T6G 1H4.
E-mail: owen.beattie@ualberta.ca

Received 11 July 1997; accepted 8 April 1998.

it takes only minutes to acquire images of each subject. Most of the landmarks could not be located on the images, so we had to mark them on each subject's face. This raised the question of what effect marking the landmarks might have on the measurement values. Consequently, although we were most interested in comparing photogrammetric measurements with those taken by caliper with the landmarks marked, we also investigated the difference between caliper-derived measurements taken with and without the landmarks marked. [Although there have been studies comparing facial measurements taken by photogrammetry and calipers, such as that by Farkas et al. (1980), they involved extracting measurements from single photographs.] The results of our study may interest anthropologists and medical clinicians who take measurements as well as those who use measurements gathered by others (e.g., normative data) when the exact techniques of data collection are unknown.

MATERIALS AND METHODS

We repeated 13 facial soft tissue measurements ten times (or as indicated in Table 1) on one adult female (Subject 1) and one adult male (Subject 2) who were also the observers in this study (Observer B and A, respectively). We used the following landmarks: exocanthion (ex), endocanthion (en), sellion (se), subnasole (sn), alare (al), cheilion (ch), stomion (sto), pogonion (pg), zygion (zy), condylion laterale (cdl), and gonion (go). The landmarks and their abbreviations were taken from Farkas (1981), except for sellion (Farkas, 1994) and gonion (Krogman, 1970). The measurements were basic ones that are used routinely for investigating facial abnormalities by one of us (J.S.B.) in clinical practice. All of the caliper-derived measurements were taken with a sliding caliper (Mitutoyo dial caliper with a smallest division of 0.01 mm), except cdl-cdl and go-go, which were measured with a spreading caliper (Abaware caliper with 1-mm divisions). The dimensions were identified as one of two types, depending on whether or not the landmarks could be marked: In the case of en and sto, neither landmark could be marked at any time, and dimensions involving either landmark were designated Type 2. All others in which both of the landmarks could be marked were classed as Type 1.

The period of data collection spanned a 22-day period, with a minimum of 24 hours and a maximum of 5 days between the repeated sets of measurements. All measurements were taken with the same protocol. For the caliper-derived measurements that were taken without marking the landmarks, the subject sat upright with his or her head held in a natural position while the 13 measurements were taken. These values were covered up before the next stage was begun. All of the landmarks of interest, except for en and sto, were then marked with a black eyeliner pencil, and the 13 measurements were taken again with calipers. Without removing the markings, the subject was seated approximately 1.5 meters in front of six Logitech FotoMan Plus cameras, each with a fixed focus lens with a focal length equivalent to a 64-mm focal length in a 35-mm camera (Logitech Inc., 1993), to which was added a $\times$1.5 enlargement Optex telephoto video lens. The cameras were arranged for frontal and oblique lateral coverage of the subject's face; all were triggered simultaneously.

The photographic images were processed by Observer B after the experiment was completed. A calibration grid was imaged at each session, and the digitized targets were used to determine each camera's perspective center coordinates, angles of rotation about the three camera axes, and the principal distance as well as the principal point coordinates of the image, one term for radial distortion of the lens, and two terms for the decentering distortion. Next, the facial landmarks were digitized, and a record was kept of the landmarks that were difficult to digitize. This information was used to decrease the weight of the digitized coordinates in a camera view when warranted by poor conditions, such as faint or no markings, an exceptionally oblique view of the marks, or shadows that obscured the marks. The appropriate camera calibration data were combined with the digitized facial landmark

data, and the three-dimensional coordinates of each landmark were determined by using the collinearity equations (Wolf, 1983). The program for this procedure automatically flagged, but did not reject, data that had large residuals of fit (greater than ±2 standard deviations), permitting reassessment. Only in cases in which a landmark clearly had been digitized incorrectly was it redigitized. Once the coordinates of each landmark had been determined, the distances between specified markings were calculated to match those measured by calipers. This program output the distances to 0.1 mm.

The caliper-derived data were analyzed after removing blunders resulting from measuring from the wrong landmarks or transcription errors (the instances in which measurement values were removed are listed in Table 1). None of the measurement values were eliminated from the photogrammetric data, because dissimilar data were flagged and reassessed before calculating distances from them. The means and standard deviations of each dimension taken by each technique were computed. Also, the probabilities associated with the t-test for paired comparisons were calculated between the means of the caliper-derived data taken with the landmarks that were unmarked and marked for each subject. The results of the statistical analyses are shown in Table 1.

RESULTS

The photogrammetric data were compared with the caliper-derived data that were taken with the landmarks marked. The photogrammetric method presented special problems for the unmarked landmarks en and sto (Type 2 measurements). For Subject 2, sto could be digitized only in the images from one measurement session, whereas, for Subject 1, it was digitized in images from all of the ten sessions. Although the en landmarks were digitized in images of both subjects from nine of the ten measurement sessions, comparison of the photogrammetric and caliper-derived (with the landmarks marked) means of dimensions, which included en, revealed that this landmark

TABLE 1. The means and standard deviations of the soft tissue facial measurements taken by photogrammetry (Photogram) and by calipers with the landmarks either unmarked (C-unmark) or marked (C-mark)[1]

Method	Type 1 mmts	Subject 1 N	Subject 1 $\bar{X}$	Subject 1 SD	Subject 2 N	Subject 2 $\bar{X}$	Subject 2 SD	Type 2 mmts	Subject 1 N	Subject 1 $\bar{X}$	Subject 1 SD	Subject 2 N	Subject 2 $\bar{X}$	Subject 2 SD
Photogram	ch-ch	10	44.4	1.7	10	55.2	1.8	en-en	9	39.2	1.7	9	49.6	3.6
C-unmark		10	43.89	1.61	9[2]	52.26**	0.75		10	30.55**	0.28	10	35.56	0.95
C-mark		9	43.43	1.72	10	53.54**	1.21		10	30.89**	0.33	10	35.95	0.64
Photogram	zy-zy	10	135.8	2.0	10	150.8	2.6	ex-en R	9	23.8	1.1	9	28.7	2.5
C-unmark		10	128.40	1.50	10	144.70	0.90		10	27.21	0.94	10	31.92**	1.29
C-mark		10	128.80	0.40	10	143.70	1.10		10	28.39	1.99	10	33.08**	1.01
Photogram	al-al	10	36.3	1.0	10	41.4	2.1	ex-en L	9	25.2	1.4	9	27.1	1.6
C-unmark		10	35.08	0.69	10	39.55	0.99		10	27.67**	0.82	10	32.39	1.32
C-mark		10	35.06	0.60	10	39.54	0.62		10	29.56**	0.76	10	33.18	1.26
Photogram	sn-pg	10	51.5	0.9	10	53.1	2.8	se-sto	10	71.0	1.3	1	80.3	NA
C-unmark		10	52.39**	1.36	9[2]	52.62	1.84		10	68.80	1.12	10	74.97*	1.77
C-mark		10	50.57**	0.69	9[2]	51.15	1.84		9[2]	69.08	1.23	10	76.97*	1.27
Photogram	se-sn	10	48.3	1.0	10	56.6	1.3	sto-pg	10	28.3	1.4	1	32.9	NA
C-unmark		8	47.91	1.35	10	54.55	2.32		10	28.93	2.41	10	33.62**	2.29
C-mark		9[2]	48.72	1.29	10	56.13	1.44		10	29.02	1.15	10	30.63**	2.60
Photogram	go-go	9	104.4	2.6	10	123.1	1.9							
C-unmark		10	93.50*	2.69	8	115.50	3.35							
C-mark		4	97.25*	0.83	8	112.88	3.18							
Photogram	cdl-cdl	10	133.5	2.9	10	151.8	2.1							
C-unmark		9	119.67**	1.63	8	137.38	2.64							
C-mark		5	123.20**	1.47	8	139.63	3.16							
Photogram	ex-ex	10	87.0	1.1	10	103.7	0.9							
C-unmark		10	82.67**	0.59	10	101.52	3.06							
C-mark		10	87.03**	0.77	10	102.83	2.05							

[1] All measurements are in mm. mmts, Measurements; N, number of measurements taken for each method; $\bar{X}$, mean of the measurements; SD, standard deviation of the measurements; NA, not available.
[2] One value was removed.
* T-test probability of $0.01 < P \leq 0.05$.
** T-test probability of $P \leq 0.01$.

was never digitized correctly. On the other hand, the similarity of the ex-ex means from measurements taken by caliper and photogrammetry indicated that these markings were digitized correctly. Of the remaining measurements, we expected the photogrammetric means to be larger than the caliper-derived means, because photogrammetry did not cause the soft tissues to be compressed. This expectation was confirmed in all of the measurements except for sto-pg and se-sn from subject 1.

The standard deviations of the photogrammetric distances (excluding those involving en) were greater than those from the caliper-derived measurements in the majority of cases for both subjects. Those dimensions in which the standard deviations were decreased were all Type 1: se-sn from Subject 1 and ex-ex, se-sn, go-go, and cdl-cdl from Subject 2.

T-tests of the means of the caliper-derived measurements taken with the landmarks unmarked and marked showed that, in ten cases out of a total of 26, the differences between the means were statistically significant at $P \leq 0.05$. Both Type 1 and Type 2 dimensions were found to be statistically significantly different, but never in both subjects' data for any one dimension. Whether this is attributable to differences between the two subjects (e.g., male vs. female, varying amounts of physiological change) or between the observers (e.g., Observer A had more experience in taking facial measurements than Observer B) is not known. It is of note that t-tests of ex-en R and ex-en L indicated that only the means from the left eye measurements from Subject 1 and from the right eye measurements from Subject 2 were statistically significantly different. Also, the means of the en-en measurements from Subject 1 were significantly different, although there were no differences in the collection procedures. Nonsignificant t-test probabilities ($P > 0.05$) were noted for both subjects for the Type 1 dimensions of zy-zy, al-al, and se-sn.

The standard deviations of the caliper-derived measurements taken with the landmarks marked showed a relative decrease in magnitude in eight cases for both subjects compared with the data taken with the landmarks unmarked. Although the majority of the decreases were of the order of 0.5 mm or less, there were some instances in which they were more substantial. Relative reductions of between 0.67 mm and 1.86 mm occurred predominantly in Type 1 measurements (zy-zy, sn-pg, and go-go from Subject 1; ex-ex and se-sn from Subject 2), with only one in the Type 2 category (sto-pg from Subject 1). Except for ex-en R from Subject 1, relative increases in the standard deviations of the caliper-derived data taken with marked landmarks, compared with the caliper-derived data taken without the landmarks marked, were less than 0.5 mm.

DISCUSSION

Comparison of the photogrammetric data with the data taken by caliper with the landmarks marked revealed a systematic difference, as was described previously by Gavan et al. (1952) and DiLiberti and Olson (1991), in addition to increased variability, which was also noted by Fraser and Pashayan (1970). The increased variability in the measurements taken by photogrammetry was probably due to three main causes. First, an oblique view of the marks presented a problem, because the center of the mark extended over several pixels, and one central pixel could not be digitized consistently. Obliquity was a common problem for the most lateral landmarks of zy, go, and cdl, but it affected all of the landmarks in accordance with each camera's position relative to each marking. Second, landmarks that were obscured by shadows were difficult to digitize. Third, when the landmarks were not marked (as in Type 2 measurements), it was hard to locate and digitize them.

On the other hand, it is possible that the standard deviations of the caliper-derived measurements taken with the landmarks marked were decreased in relation to the photogrammetric data, because successive values were influenced by previous ones, despite each observer believing that they forgot the results shortly after taking each measurement (Fleiss, 1986). Photogrammetry essentially eliminated this problem, because the distances were produced by computer. The standard deviations that were calculated from the data gathered by using

this method probably show the most realistic amounts of variability in repeated measurements.

The t-test results from the two sets of caliper-derived data suggest two possibilities: There is naturally a great amount of variability in the measurements, or there is a systematic difference between the data gathered by the two techniques. The first interpretation is bolstered by the statistically significant t-test result between the means for the dimension en-en from Subject 1. However, the measurement values taken with the landmarks marked have larger means in ten cases for the data on Subject 1 and in eight cases for Subject 2 as well as decreased standard deviations in eight cases for the data from both subjects, which strongly suggests that there is a systematic difference between the two techniques.

Marking the landmarks had a positive effect in controlling two sources of variability in the caliper-derived measurements: different amounts of caliper pressure on highly compressible areas and accidental slippage of the calipers off of the landmarks. The first case is demonstrated by the dimension al-al. The means of the caliper-derived data taken with the landmarks unmarked and marked were nearly identical for each subject, but the standard deviations of the data taken with the landmarks marked were decreased. Because this was a measurement in which both landmarks were visible without the observer shifting position, there was no great advantage conferred by marking the landmarks other than to allow the observer to focus on the amount of pressure being exerted on the skin by the calipers. The second advantage is demonstrated by the dimension go-go. This was a particularly difficult measurement to take, because each landmark could be palpated and viewed by the observer only by shifting to one side. With the landmarks marked, any slippage of the calipers could be corrected easily. Neither of these advantages, however, was able to negate the basic problem of locating troublesome landmarks, such as cdl. Also, this procedure did not appear to be beneficial when the landmark was highly mobile, as in the case of ch.

Marking the landmarks in the sensitive areas of the eyes (i.e., ex) did not appear to improve consistently our ability to measure these features. However, when each subject's means for the right and left ex-en measurements and the mean distance between en-en were added together, the absolute differences between the estimated ex-ex lengths and those actually measured were the greatest when all of the landmarks were unmarked: 2.76 mm (vs. 1.81 mm when ex was marked) for the data from Subject 1 and 1.65 (vs. 0.62 mm when ex was marked) for the data from Subject 2. Marking the landmarks before taking the measurements with calipers helped make multiple measurements taken from them more uniform, as suggested by Farkas (1981, 1994) and Farkas and Deutsch (1996). When the same computations were done for the photogrammetric means, the data from Subject 1 showed the lowest difference (1.2 mm), whereas the data from Subject 2 demonstrated the highest difference (1.7 mm).

Our investigation found that there were systematic differences between photogrammetric measurements and caliper-derived measurements (with the landmarks marked) as well as between caliper-derived data taken with the landmarks marked and unmarked. Researchers and clinicians who rely on measurements to objectify their diagnoses or classifications of subjects should be aware that small (i.e., marking or not marking the landmarks) and large (photogrammetric or caliper-derived measurements) differences in measurement techniques may result in statistically significant findings, even in data taken from the same subject.

ACKNOWLEDGMENTS

We thank the anonymous reviewers for their suggestions on how to improve our paper. Thanks also to the Department of Anthropology, University of Alberta for the calipers used in this study, and to the University of Alberta Hospital for the use of the room in which we collected the measurement data.

LITERATURE CITED

Allanson JE, O'Hara P, Farkas LG, and Nair RC (1993) Anthropometric craniofacial pattern profiles in Down syndrome. Am. J. Med. Genet. *47*:748–752.

DiLiberti JH, and Olson DP (1991) Photogrammetric evaluation in clinical genetics: Theoretical considerations and experimental results. Am. J. Med. Genet. *39:*161–166.

Farkas LG (1981) Anthropometry of the Head and Face in Medicine. New York: Elsevier.

Farkas LG (1994) Examination. In LG Farkas (ed.): Anthropometry of the Head and Face, 2nd Ed. New York: Raven Press, pp. 3–56.

Farkas LG, and Deutsch CK (1996) Anthropometric determination of craniofacial morphology. Am. J. Med. Genet. *65:*1–4.

Farkas LG, Bryson W, and Klotz J (1980) Is photogrammetry of the face reliable? Plas. Reconst. Surg. *66:*346–355.

Fleiss JL (1986) The Design and Analysis of Clinical Experiments. New York: John Wiley and Sons.

Fraser FC, and Pashayan H (1970) Relation of face shape to susceptibility to congenital cleft lip. J. Med. Genet. *7:*112–117.

Gavan JA, Washburn SL, and Lewis PH (1952) Photography: An anthropometric tool. Am. J. Phys. Anthropol. *10:*331–351.

Krogman WM (1970) Growth of head, face, trunk, and limbs in Philadelphia white and Negro children of elementary and high school age. Monogr. Soc. Res. Child Dev. S 136 *35(3):*1–80.

Logitech Inc. (1993) Logitech FotomanPlus User's Guide. Freemont, CA: Logitech, Inc.

Wolf PR (1983) Elements of Photogrammetry With Air Photo Interpretation and Remote Sensing, 2nd Ed. New York: McGraw-Hill, Inc.